DOI: 10.1002/anie.200((will be filled in by the editorial staff))

Probing the Molecular Recognition of a DNA·RNA Hybrid Duplex**

Richard T. Wheelhouse, Nichola C. Garbett, Niklaas J. Buurma and Jonathan B. Chaires.*

DNA·RNA hybrid duplexes are the nucleic acid structures least explored as targets in drug discovery. There is a dearth of model ligands displaying convincing structure or sequence selectivity; the secondary structures of the hybrid nucleic acid hosts are variable and complex[1, 2] Ligands selective for hybrid duplexes have potential therapeutic applications as telomerase and RNaseH inhibitors.

Assays for investigating small molecule interactions with DNA·RNA hybrid duplexes have only recently been developed and described.[3] Herein their application to the discovery of a small molecule that specifically recognizes poly(dA)·poly(rU) is reported. An unexpected binding mode was uncovered, establishing a principle that extrapolations from pure DNA or RNA binding are invalid when considering DNA·RNA hybrid duplexes.

Competition dialysis (Fig. 1) showed that the non-classical intercalator **1** bound preferentially to quadruplex and triplex DNA structures and to the hybrid duplex poly(dA)·poly(rU). In duplex binding, strong preferences for both backbone and base orientation were evident: association with the poly(dA)·poly(rU) hybrid duplex had ~20-fold preference over the equivalent RNA, 3-fold over DNA, and 7-fold over the alternative poly(rA)·poly(dT) hybrid.

Relative binding preferences for duplex structures were confirmed in a melting of mixtures assay in which UV melting profiles were obtained for a mixture of all four A·T(U) duplexes in the absence and presence of low concentrations of ligand, Fig. 2.[3, 4] Significant shifting of the melting transition of the poly(dA)·poly(rU) hybrid was accompanied by a lesser shift in the DNA duplex melting curve at higher ligand loading.

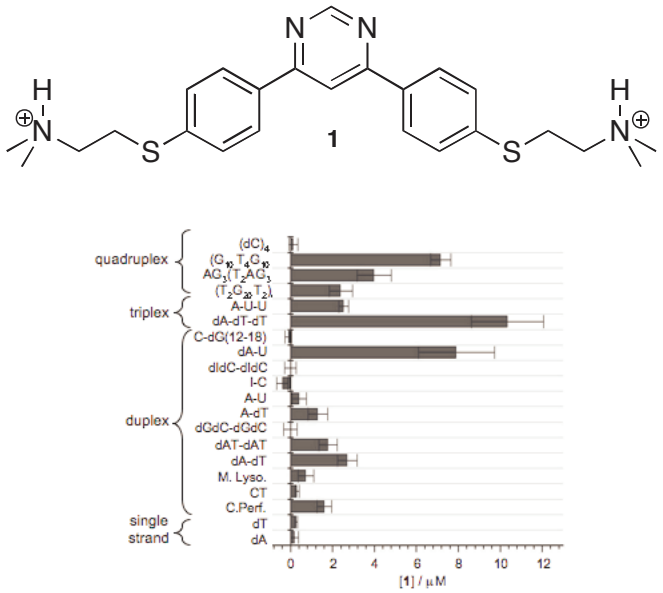


Figure 1. Competition dialysis data for compound **1**, all data are the mean±SD of 3 separate determinations.

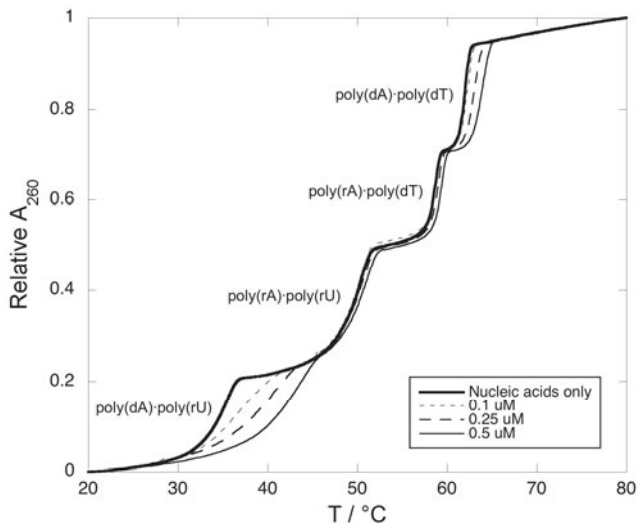


Figure 2. Melting of mixtures data for the four A·T(U) duplexes (each 10 M(bp)) in the presence of increasing concentrations of ligand **1**.

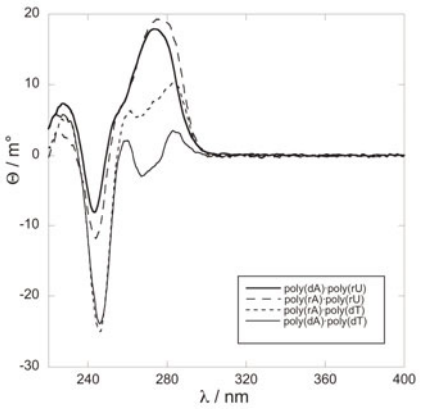
[*] R.T. Wheelhouse
School of Pharmacy
University of Bradford
Bradford, BD7 1DP, UK
E-mail: r.t.wheelhouse@brad.ac.uk

J.B. Chaires and N.C. Garbett,
James Graham Brown Cancer Center
University of Louisville
Clinical and Translational Research Building
505 S Hancock Street
Louisville, KY 40202, USA

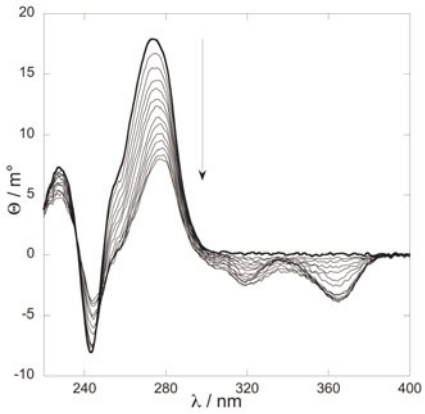
N.J. Buurma
Physical Organic Chemistry Centre,
School of Chemistry,
Cardiff University,
Main Building, Park Place,
Cardiff, CF10 3AT, UK

[**] We thank the British Association for Cancer Research for a Mid-career Fellowship (RTW); NIH Grant Number GM077422 (JBC).

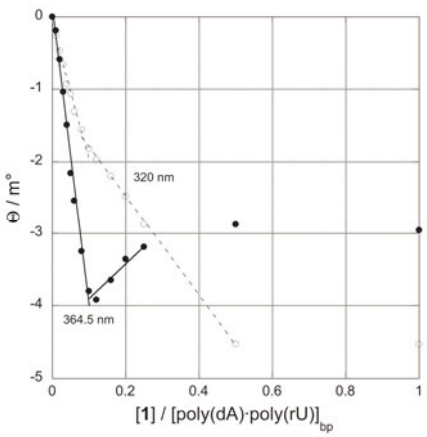
Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.



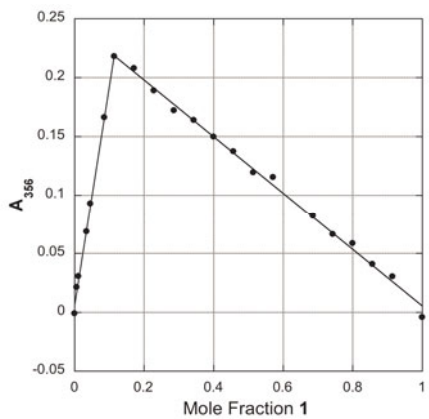
a)



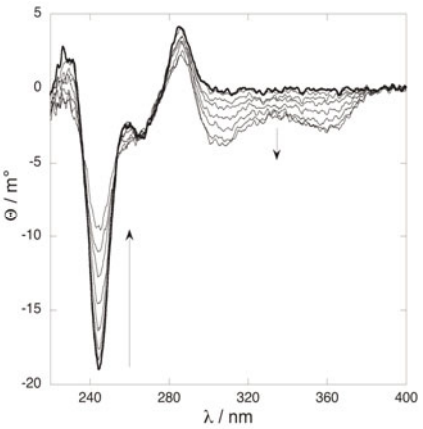
b)



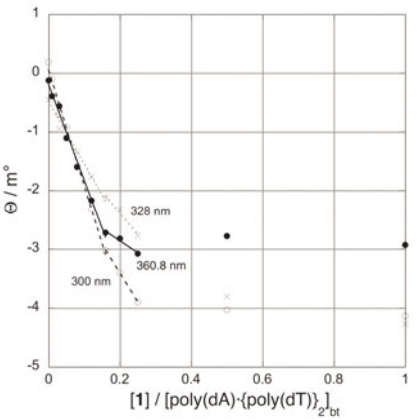
c)



d)



e)



f)

Figure 3. a) Reference CD spectra of the four A·T(U) duplexes; b) CD titration of ligand **1** into poly(dA)·poly(rU) duplex; c) Induced CD as a function of the [**1**] / $[poly(dA)\cdot poly(rU)]_{bp}$ ratio; d) Job plot for mixtures of compound **1** with poly(dA)·poly(rU) at a constant 100 M. e) CD titration of ligand **1** poly(dA)·$[poly(dT)]_2$ triplex, the spectrum of the free nucleic acid is shown as a heavy line; f) induced CD as a function of the [**1**] / $[poly(dA)\cdot\{poly(dT)\}_2]_{bt}$ ratio.

CD experiments (Fig. 3a–c) further elucidated the binding interaction. As ligand **1** was titrated into the hybrid solution, the intensity of nucleic acid ellipticity (244 and 275 nm) decreased but the envelope of the spectrum retained its starting shape: indicative of reduced chirality in the oligonucleotide (unwinding) due to intercalation, without alteration of the global conformation. Negative induced CD bands observed for the ligand (320 and 364.5 nm, Figs 3b,c) were also consistent with intercalation.[5] Intercalation was confirmed by a preliminary analytical ultracentrifugation (AUC) experiment in which the sedimentation coefficient decreased upon ligand loading in a manner consistent with unwinding and lengthening of the duplex (Fig S2). Plots of induced CD intensity vs. ligand:base pair ratio (Fig. 3c) had an inflection at a ratio of 1:10 (at high ligand concentrations precipitation of the sample occurred). The 10 bp per ligand binding stoichiometry was confirmed by a Job plot (Fig. 3d).

CD titration spectra using poly(dA)·$[poly(dT)]_2$ triplex DNA, Fig. 3e, reflected those obtained for the hybrid duplex, showing unwinding of the triplex and negative induced CD in the ligand (300, 360.8, 328 nm). The inflection in the induced CD plots (Fig. 3f) occurred at a mixing ratio of 0.15, i.e. ~7 bt per ligand.

Isothermal titration calorimetry (ITC) probed thermodynamic details of the binding interaction. The thermogram for ligand dilution indicated disaggregation of the ligand; data analysis using an isosdesmic self-aggregation model yielded $K_{agg} = 4.1\pm0.7\times10^2$ M^{-1} and $\Delta H_{agg} = -7.3\pm0.6$ kcal mol^{-1} (Fig. S3, S4). The thermogram for ligand–hybrid interaction indicated (at least) two types of binding site (Fig. 4). ITC data were analyzed using a model describing two consecutive binding events and a model describing two independent binding events; both models incorporated simultaneous ligand self aggregation (Scheme S1). Strong parameter correlation meant that the two types of site could not be quantified independently in a statistically-meaningful way. Nevertheless, analysis of parameter correlation plots concluded that the two binding events are a major binding site of 10 bp in combination with a minor binding event involving ligand association with (potentially-fraying) ends and gaps (supporting information). The overall binding stoichiometry is consistent with other experimental data for this system and simulated data based on binding parameters from ITC reproduced UV titrations well (Fig. S10).

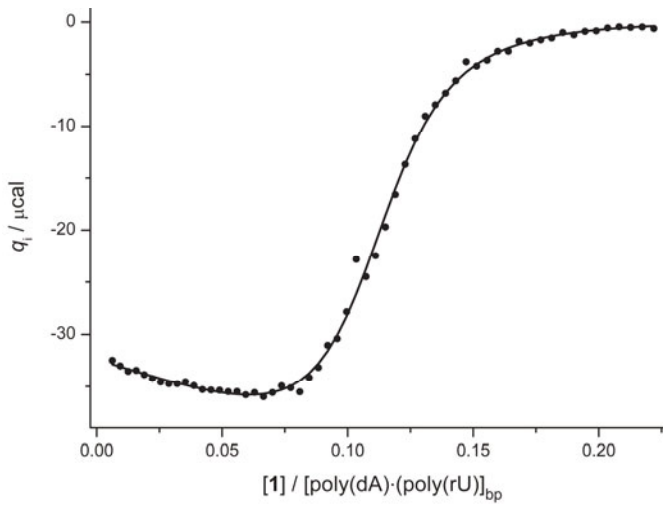


Figure 4. Integrated heat effects for titration of ligand **1** into poly(dA)·poly(rU) fitted using a model involving a major binding site in combination with a minor binding event (stoichiometry fixed at 1:0.005).

In conclusion, biarylpyrimidine **1** shows a strong preference for binding the poly(dA)·poly(rU) hybrid over all other duplex nucleic acid structures and as a DNA·RNA hybrid binder, inhibited RNaseH with $IC_{50} = 29.5 \pm 0.2$ M (Fig. 5).

All experimental data indicated that each ligand occupies a 10 bp site; CD and AUC data unambiguously showed nucleic acid intercalation. The non-classical intercalator has flexibility of torsion angles between the three aromatic rings, this may contribute to the binding preference for the A-like hybrid as the ligand can accommodate closely to propeller-twisted base pairs.[6] Ligand **1** also showed binding preferences for triplex and quadruplex nucleic acids. Moreover, triplex poly(dA)·[poly(dT)]$_2$ also exists in an A-like helix. It seems, therefore, that the binding preferences discovered are selection for the global A-like structure. However, ligand **1** does not bind to the other A-form structure in Fig. 1, the RNA duplex poly(rA)·poly(rU), an observation in contrast to major groove binding aminoglycosides where binding to equivalent RNA duplexes was strong.[7] An explanation may lie in the grooves and the dimensions of the ligand. The ligand possesses two cationic side chains and it may be postulated that one lies in the minor groove. Although the 2'-OH and associated solvation waters of a single RNA strand offer potential hydrogen bonding sites to the terminal ammonium functionality, these groups occlude the minor groove of the pure RNA duplex. Such steric constraints would also account for the reduced affinity of bulkier analogs of ligand **1** (Fig. S11). Indeed the hybrid minor groove appears inaccessible to all but the smallest groups: pentamidine, berenil, Hoechst 33258 and DAPI all failed to stabilize either of the hybrid duplexes in the melting of mixtures assay.[3]

Binding rules for DNA·RNA duplexes are ill-defined and this study shows that naïve extrapolation from duplex binding models is inappropriate. The structural basis for the ~10 bp major binding site size is not yet clear but is equivalent to ~1 ligand per helical turn. The intercalator ethidium has been shown to exclude 3 or 7 bp when bound to the poly(dA)·poly(rU) hybrid.[8, 9] and precedent for long-range transmission of binding effects over tens to hundreds of bp exists in the daunorubicin-driven Z→B conversion of [poly(dGdC)]$_2$.[10]

DNA·RNA hybrids and DNA triplexes both adopt structures intermediate between classic B-form DNA and A-form RNA. The correlation between binding data for ligand **1** and its analogs to these two structures is striking (Figs. 3 and S12) and demonstrates that the search for structure and sequence selective ligands for DNA·RNA hybrid duplexes should start with re-evaluation of DNA triplex-targeting compounds.

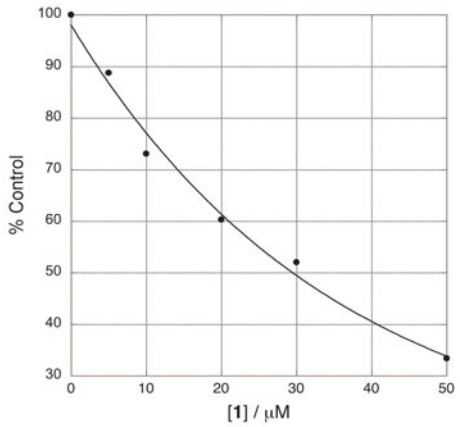


Figure 5. Inhibition of the RNaseH digestion of poly(dA)·poly(rU) by compound **1**; $IC_{50} = 29.5 \pm 0.2$ M.

Experimental Section

Compound **1** was synthesized as previously described.[11, 12] Oligonucleotides for CD and calorimetry were obtained from Midland Certified Reagent Co, Midland, TX (poly(dA), lot 111997, 126–1200 nucleotides; poly(rU) lot 102197, 500–2600 nucleotides). All other oligonucleotides were purchased from Sigma, Milwaukee, WI, USA or Poole, UK.

UV melting experiments were performed on a Cary 400Bio spectrophotometer equipped with Peltier. CD spectra were recorded on a Jasco J-810 spectropolarimeter with Peltier. ITC was performed on a Microcal VP-ITC calorimeter. Other spectra were recorded on a Jasco V-550 spectrophotometer and Tecan Safire^2 microplate reader. Competition dialysis,[3, 13-16] melting of mixtures [3, 4] and RNaseH assays[3, 4] followed published protocols.

Job plots: the protocol followed that set out by Jenkins[17] but detected by UV. Mixtures of **1** and poly(dA)·poly(rU) to a summed concentration of 100 M in BPES pH 6.00 were set up in duplicate on a 96 well microtitre plate. Reference solutions had buffer in place of nucleic acid. Spectra were recorded from 230-450 nm. Absorbance data for the drug maximum at 356 nm were extracted and the free drug absorbance subtracted from that of the drug-nucleic acid complex before plotting A_{356} vs. mole fraction of ligand

CD Titrations: nucleic acid solutions in BPES pH 6.00 (1 ml) were prepared to have A_{257} ~0.8 and placed in a 1 cm pathlength quartz cuvette equipped with stirrer. Concentrated ligand solutions were prepared in the same buffer to minimize volume changes during titration. Aliquots were added until precipitation of the ligand-nucleic acid complex was seen. Raw data were corrected for dilution before plotting and analysis.

ITC: all experiments were performed in BPES pH 6.00. Oligonucleotides were dissolved in buffer, dialyzed against buffer (3 changes of solution) and the final dialysate used as solvent for all reactions and dilution of reagents. In a typical experiment, a solution of poly(dA)·poly(rU) (0.5 mM(bp)) was loaded into the calorimeter cell and ligand **1** (0.25 mM) loaded into the syringe; aliquots (10 l) were added at intervals of 360 s. To determine the heats of dilution, a solution of **1** (5 mM) was titrated (15 l aliquots, 360 s) into buffer. Other control titrations of buffer into nucleic acid and buffer into buffer were also performed. Data were treated in Origin (Microcal, Inc) to generate integrated heat effects per injection (Δh). These were analyzed using IC ITC following published procedures.[18, 19] Ligand self-aggregation parameters were determined from ligand dilution experiments and held constant in subsequent analyses of binding data.

Received: ((will be filled in by the editorial staff))
Published online on ((will be filled in by the editorial staff))

Keywords: DNA·RNA hybrid· calorimetry· RNaseH· biophysics· pyrimidine

[1] N. N. Shaw, D. P. Arya, *Biochimie* **2008**, *90*, 1026.
[2] J. I. Gyi, G. L. Conn, A. N. Lane, T. Brown, *Biochemistry* **1996**, *35*, 12538.
[3] R. T. Wheelhouse, J. B. Chaires, in *Drug-DNA Interaction Protocols*, 2nd ed. (Ed.: K. R. Fox), Springer, **2010**.
[4] X. C. Shi, J. B. Chaires, *Nucleic Acids Res.* **2006**, *34*, e14.
[5] N. C. Garbett, P. A. Ragazzon, J. B. Chaires, *Nature Protoc.* **2007**, *2*, 3166.
[6] W. D. Wilson, F. A. Tanious, R. A. Watson, H. J. Barton, A. Strekowska, D. W. Boykin, L. Strekowski, *Biochemistry* **1989**, *28*, 1984.
[7] C. M. Barbieri, T. K. Li, S. Guo, G. Wang, A. J. Shallop, W. Pan, G. Yang, B. L. Gaffney, R. A. Jones, D. S. Pilch, *J. Am. Chem. Soc.* **2003**, *125*, 6469.
[8] E. A. Lehrman, D. M. Crothers, *Nucleic Acids Res.* **1977**, *4*, 1381.
[9] N. N. Shaw, H. Xi, D. P. Arya, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 4142.
[10] X. Qu, J. O. Trent, I. Fokt, W. Priebe, J. B. Chaires, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 12032.
[11] R. T. Wheelhouse, S. A. Jennings, V. A. Phillips, D. Pletsas, P. M. Murphy, N. C. Garbett, J. B. Chaires, T. C. Jenkins, *J. Med. Chem.* **2006**, *49*, 5187.
[12] P. M. Murphy, V. A. Phillips, S. A. Jennings, N. C. Garbett, J. B. Chaires, T. C. Jenkins, R. T. Wheelhouse, *Chem. Commun.* **2003**, 1160.
[13] J. S. Ren, J. B. Chaires, *Biochemistry* **1999**, *38*, 16067.
[14] J. S. Ren, J. B. Chaires, *J. Am. Chem. Soc.* **2000**, *122*, 424.
[15] J. S. Ren, J. B. Chaires, *Methods Enzymol.* **2001**, *340*, 99.
[16] J. B. Chaires, *Current Med. Chem. Anticancer Agents* **2005**, *5*, 339.
[17] T. C. Jenkins, in *Drug-DNA Interaction Protocols, Vol. 90* (Ed.: K. R. Fox), Humana Press Inc., Totowa, N.J., U.S.A., **1997**, pp. 195.
[18] N. J. Buurma, I. Haq, *J. Mol. Biol.* **2008**, *381*, 607.
[19] N. J. Buurma, I. Haq, *Methods* **2007**, *42*, 162.

Entry for the Table of Contents (Please choose one layout)

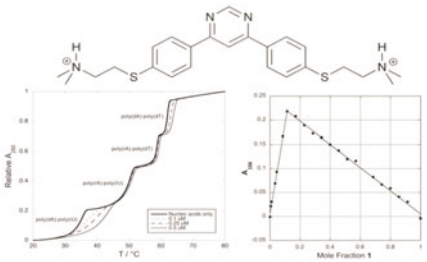
Layout 1:

Targeting a DNA·RNA Hybrid

Richard T. Wheelhouse,* Nichola C. Garbett, Niklaas J. Buurma and Jonathan B. Chaires.

_____ **Page – Page**

Probing the Molecular Recognition of a DNA·RNA Hybrid Duplex



Targeting a DNA·RNA Hybrid: A birarylpyrimidine ligand shows a marked structure and sequence selectivity for the poly(dA)·poly(rU) hybrid duplex. Intriguingly the binding event involves intercalation with a stoichiometry of 10 base pairs per ligand.

Probing the Molecular Recognition of a DNA·RNA Hybrid Duplex

Richard T. Wheelhouse, Nichola C. Garbett, Niklaas J. Buurma and Jonathan B. Chaires.*

Supporting Information

Page

Circular Dichroism

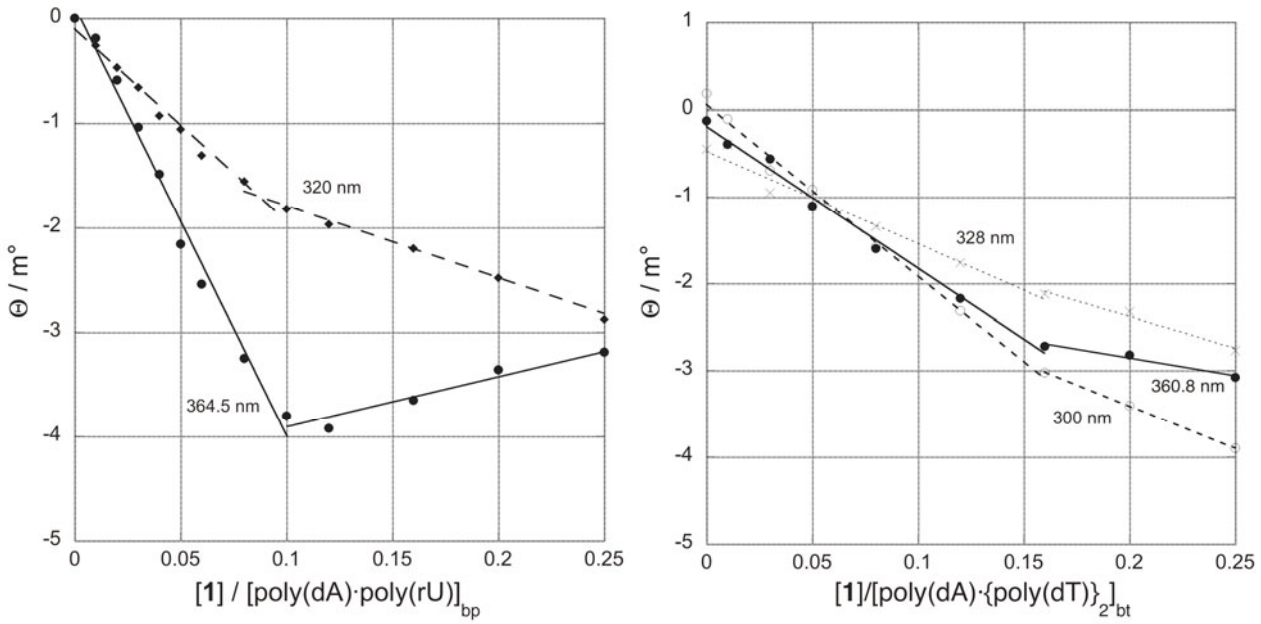


Figure S1. Expansions of Figures 3c, f.

Analytical Ultracentrifugation (AUC)

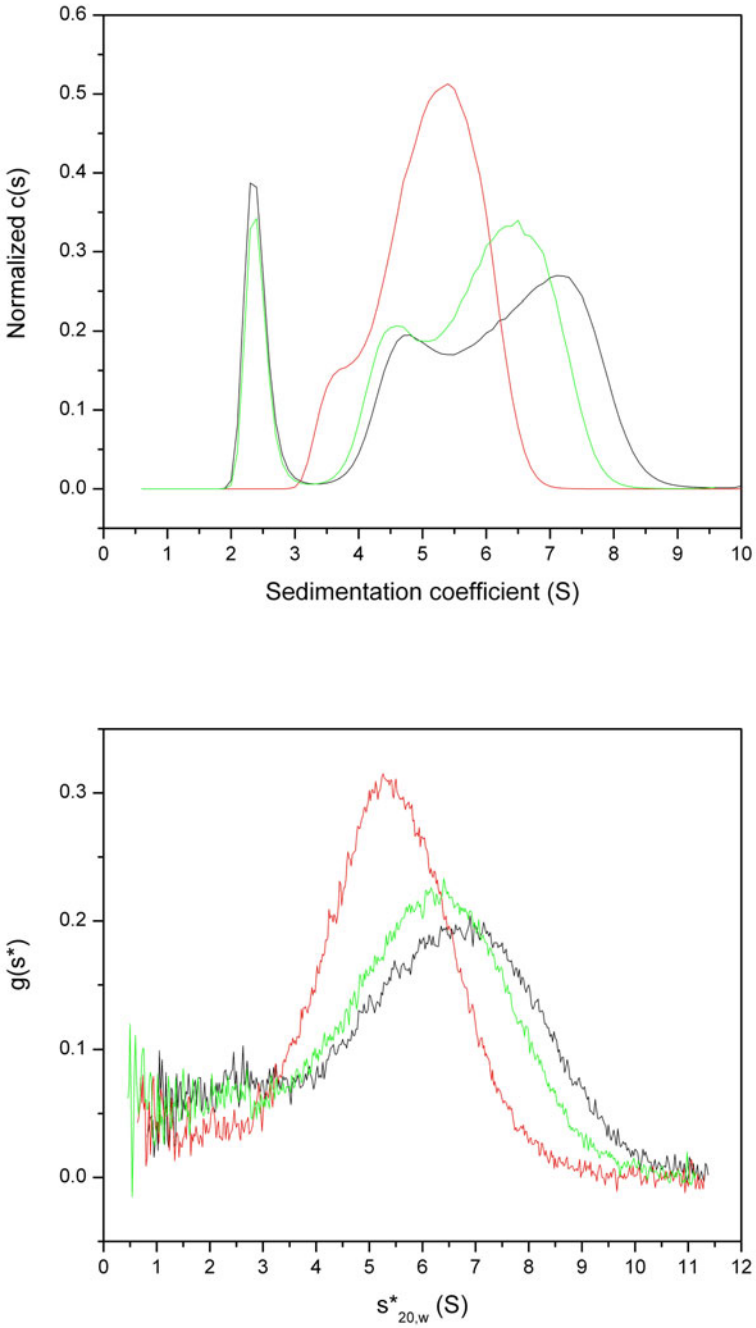


Figure S2. Sedimentation velocity AUC data. Solutions of poly(dA)·poly(rU) (0.7 ml, A_{257} =0.5, 0.0385 mM(bp), BPES pH 6.00) were prepared alone and in the presence of ethidium (0.0385 mM) and ligand **1** (0.0077 mM). Cell 1 (black), hybrid; Cell 3 (green), hybrid+**1**; Cell 2 (red), hybrid + ethidium.

AUC Method

Sedimentation velocity experiments were performed at a temperature of 20 ^{o}C and a rotor speed of 50,000 rpm using a Beckman Optima XL-A analytical ultracentrifuge equipped with absorbance optics and an An60Ti rotor. Following loading and before data collection, samples were allowed to equilibrate for 1 h after vacuum and temperature had been established. Data were collected at 260 nm as a function of radial position. Each centrifuge cell was scanned sequentially with zero time delay between scans until no further sedimentation was observed. Primary sedimentation data were transferred to the

programs Sedfit and DCDT+ for analysis (1-5). Solution density and viscosity were calculated from buffer composition as 1.00712 g/mL and 1.0225 cP, respectively, using the program Sednterp (6). A value of 0.55 mL/g was assumed for the partial specific volume (7, 8). Data were analyzed using Sedfit employing a continuous c(s) model with a range of 0.5 to 10 S and a confidence level of 0.68 (1 standard deviation). Fitting was performed using alternating rounds of the simplex and Marquardt-Levenberg algorithms until there was no further decrease in rmsd. Data were transformed into a differential apparent sedimentation coefficient distribution, g(s*), using DCDT+ and evaluated without model fitting. Data in the form of c(s) and g(s*) distributions were exported to Origin v7.0 (OriginLab Corporation, Northampton, MA) for graphing.

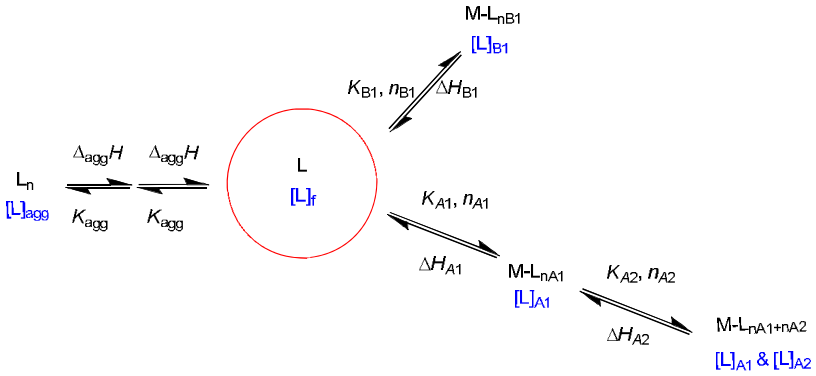
References

1. Schuck, P. (2008) SEDFIT (version 11.3b). National Institutes of Health, Bethesda, MD. Available from http://www.analyticalultracentrifugation.com/download.htm.
2. Schuck, P. (2000) Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. *Biophys. J.* **78,** 1606-19.
3. Philo, J. S. (2007) DCDT+ (version 2.1.0.28333). John S. Philo, Thousand Oaks, CA. Available from http://www.jphilo.mailway.com/dcdt+.htm.
4. Philo, J. S. (2000) A Method for Directly Fitting the Time Derivative of Sedimentation Velocity Data and an Alternative Algorithm for Calculating Sedimentation Coefficient Distribution Functions. *Anal. Biochem.* **279,** 151-63.
5. Philo, J. (2006) Improved methods for fitting sedimentation coefficient distributions derived by time-derivative techniques. *Anal. Biochem.* **354,** 238-46.
6. Hayes, D. B., Laue, T., and Philo, J. (2006) Sedimentation Interpretation Program (version 1.09). University of New Hampshire, Durham, NH. Available from http://www.jphilo.mailway.com/download.htm.
7. Bloomfield, V. A., Crothers, D. M., and Tinoco, I., Jr. (2000) *Nucleic Acids: Structures, Properties and Functions*. University Science Books, Sausalito, CA.
8. Hellman, L. M., Rodgers, D. W., and Fried, M. G. (2009) Phenomenological partial-specific volumes for G-Quadruplex DNAs. *Eur. Biophys. J.***,** Epub 24 February 2009.

ITC data analysis

Binding models

All ITC data was analysed using IC ITC. [1, 2] Parameters were defined as in Scheme S1 where L refers to ligand **1** and M (macromolecule) refers to the hybrid.



Scheme S1

Data analysis: ligand 1 self aggregation

Ligand dilution experiments were analysed globally using an isodesmic self aggregation model (stepwise aggregation model) characterised by a single equilibrium constant and a single interaction enthalpy for the subsequent addition processes of a monomer to an existing aggregate, *viz.* K_{agg} and ΔH_{agg}. An isosdesmic self aggregation model was selected over the mathematically equivalent dimerization model because the structure of ligand **1** does not suggest that self aggregation stops at dimer formation. The isosdesmic self aggregation model accurately reproduces the experimental data (Figure S3).

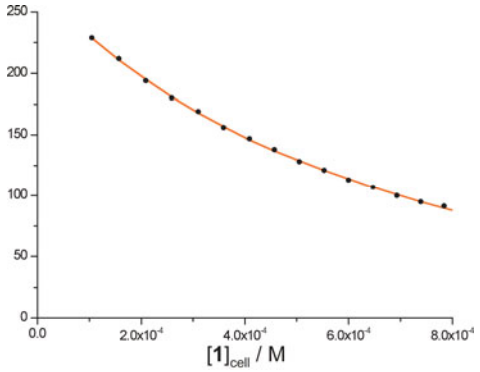


Figure S3: Integrated heat effects for dilution of a 5.0 mM ligand **1** solution in BPES, pH 6.0 at 20 °C ( ) analysed globally in terms of an isosdesmic self aggregation model (solid red line).

Error margins were determined as reported previously[2] (Figure S4).

Figure S4: Normalised Σdev^2/dof as a function of optimizable variable values.

Data analysis: nucleic acid binding

The data were initially analysed in terms of a model involving two different sets of independent binding sites, i.e. in terms of equilibria A1 and B1. The analysis resulted in fits that reproduce the experimental data well. However, analysis of the error margins showed that errors are typically large. Analysis of $\Sigma dev^2/dof$ as a function of the values of the different variables shows why the error margins on the individual variables are large. Considerable parameter correlation allows various combinations of the different parameters to reproduce the experimental data (statistically) well. The most obvious example of this parameter correlation is provided by n_{A1} and n_{B1} (Figure S5)

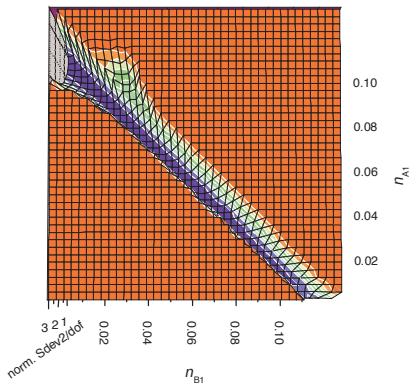


Figure S5: Normalised $\Sigma dev^2/dof$ as a function of values of n_{A1} and n_{B1}.

Figure S5 shows that the sum $n_{A1}+n_{B1}$ is well defined at 0.11. The individual parameters n_{A1} and n_{B1}, however, can not be defined using the experimental data.

Nevertheless, the observation that the SVD of the UV visible data indicated just two species, together with the Job plot showing a clean single peak, suggested that one of the following is the case: 1) there are two types of binding site which both contribute significantly to the total binding stoichiometry but have very similar (UV-visible and CD) properties and we could therefore expect similar binding constants and enthalpies of interaction or 2) there are two different binding sites but one of these contributes significantly more towards the total binding stoichiometry than the other, resulting in the low stoichiometry binding site being practically invisible by the various spectroscopic techniques.

1. Two types of similar binding site

The extreme, but representative, example of this situation involves two very similar binding sites with identical binding site sizes. For the present case, this corresponds to stoichiometries of 0.056 (*vide supra*). This binding site size corresponds to approximately two helical turns per ligand, which seems rather considerable for ligand **1**. A possible interpretation involves the binding of one ligand molecule to a site equivalent to one helical turn. This first binding event then affects one further helical turn of neighbouring bp to form a slightly different second binding site. Effectively, this corresponds to a description involving slight anti-cooperativity in ligand binding. The appropriate model to analyse this type of data would be the K_{A1}-K_{A2} model where the first binding event "creates" a second binding site. An unrestricted fit of the data in terms of the K_{A1}-K_{A2} model shows strong parameter correlation between n_{A1} and n_{A2} (Figure S6).

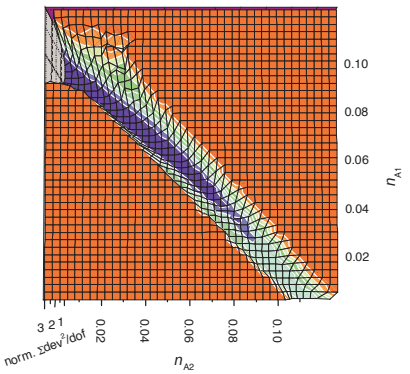


Figure S6: Normalised $\Sigma dev^2/dof$ as a function of value of n_{A1} and n_{A2}.

In order to model a system with two binding sites of (nearly) identical sizes, n_{A1} was restricted to 0.0506 in a subsequent analysis, yielding optimised parameters as illustrated in Figure S7.

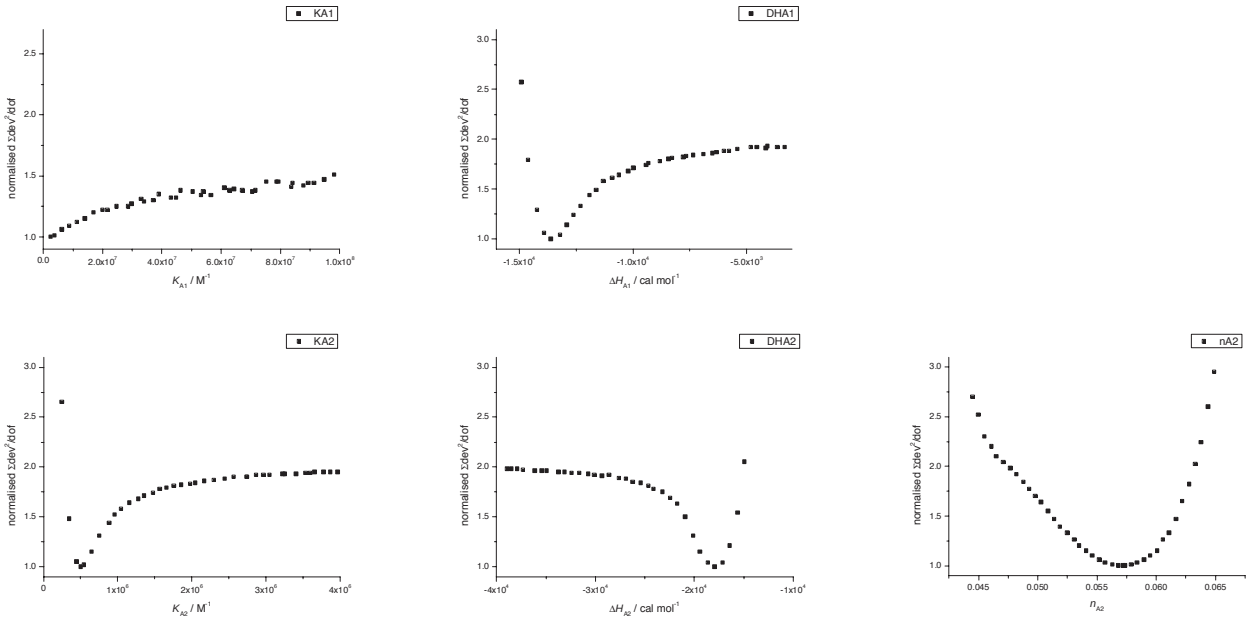


Figure S7: Normalised $\Sigma dev^2/dof$ as a function of variable value for K_{A1}, ΔH_{A1}, K_{A2} and ΔH_{A2}, n_{A2} for n_{A1} restricted to 0.056.

Figure S7 shows that ΔH_{A1}, K_{A2}, ΔH_{A2}, and n_{A2} are now reasonably well defined (although, admittedly, the plots of the normalised $\Sigma dev^2/dof$ level off just *below* 2). The best fit parameters are $K_{A1} = 2.5\times10^6$ M^{-1}, $K_{A2} = 5.1\times10^5$ M^{-1}, $\Delta H_{A1} = -1.36\times10^4$ cal mol^{-1}, $\Delta H_{A2} = -1.79\times10^4$ cal mol^{-1}, $n_{A2} = 5.7\times10^{-2}$.

2. One major binding site and one minor binding site

The alternative interpretation of the ITC data involves one binding site with a size corresponding to one helical turn and a second binding site with a stoichiometry significantly less than one per helical turn. The only logical binding site type presenting binding sites with an apparent site size significantly larger than a single helical turn are sites corresponding to strand-ends. To decide which of the two binding events, A1 or B1, is most likely to correspond to ligand **1** binding to (potentially fraying) ends and/or gaps, we analysed the correlation between stoichiometry and interaction enthalpy for both binding events (Figure S8).

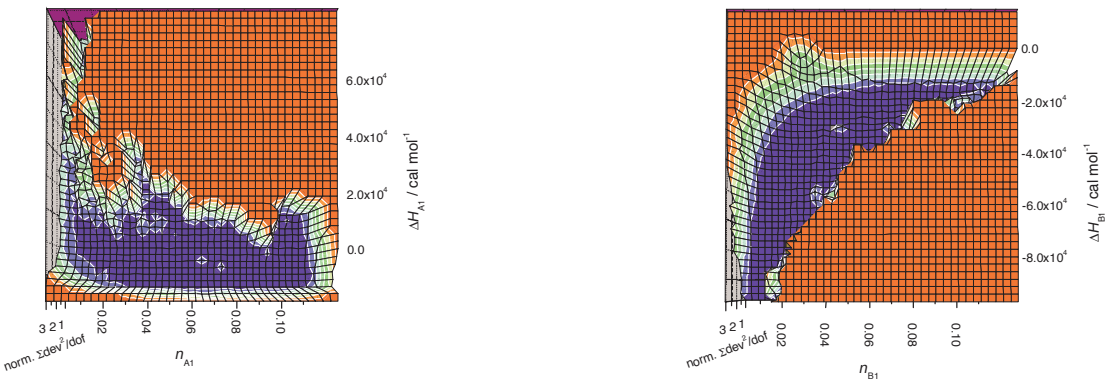


Figure S8: Normalised $\Sigma dev^2/dof$ as a function of values of n_{A1} and ΔH_{A1} (*left*) and as a function of n_{B1} and ΔH_{B1} (*right*).

Figure S8 (left) shows that the "valley" defined by the normalised $\Sigma dev^2/dof$ as a function of values of n_{A1} and ΔH_{A1} has a smooth edge defining the minimum value for ΔH_{A1} for each value of n_{A1}. The edge defining the maximum value of ΔH_{A1} for each value of n_{A1} is ragged, however, indicating that this region of parameter space was not sampled sufficiently. This was confirmed by further independent extensive simulated annealing runs which resulted in further graphs of the type of Figure S8 (left). For these graphs, the lower edges of the valleys (defining the minimum values for ΔH_{A1}) were identical to that shown in Figure 8 (left) while the edges defining the maximum values for ΔH_{A1} were not reproducible. This indicates that only the minimum values for ΔH_{A1} are well-defined. Similarly, for binding event B1, only the maximum values for ΔH_{B1} are well-defined.

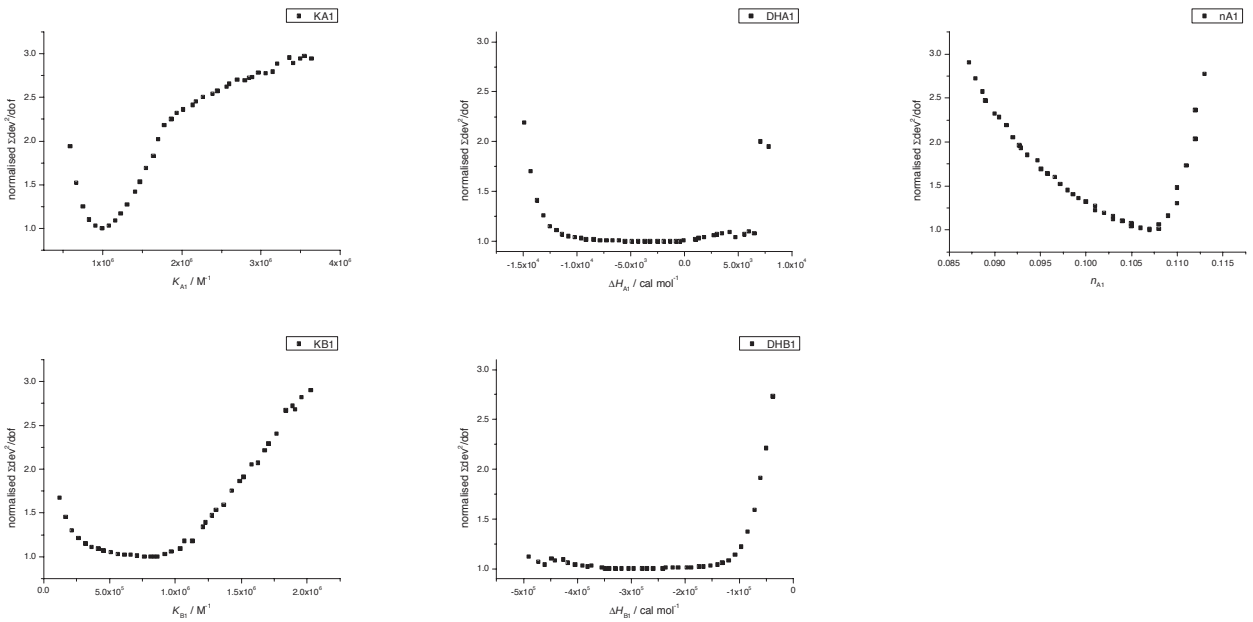


Figure S9: Normalised $\Sigma dev^2/dof$ as a function of variable value for K_{A1}, ΔH_{A1}, n_{A1}, K_{B1} and ΔH_{B1} for n_{B1} restricted to 0.005.

Figure S9 shows that K_{A1} and K_{B1} are reasonably well defined but this is not the case for the interaction enthalpies (in fact, ΔH_{B1} is unrealistically negative, but this is strongly dependent on the choice for n_{B1} as shown by Figure S8). The best fit parameters are $K_{A1} = 9.97\times10^5$ M^{-1}, $\Delta H_{A1} = -4.13\times10^2$ cal mol^{-1}, $n_{A1} = 1.07\times10^{-1}$, $K_{B1} = 8.42\times10^5$ M^{-1}, $\Delta H_{B1} = -3.49\times10^5$ cal mol^{-1}.

Nevertheless, Figure S8 shows that the behaviour of the binding enthalpy at low stoichiometry for the two binding events is rather different. For low values of n_{A1}, (the minimum value for) ΔH_{A1} becomes increasingly less exothermic and eventually even endothermic. Event A1 has been defined to be the first binding event, i.e. it corresponds to the highest binding constant, and would therefore have to be significantly, if not completely entropy driven. Binding event A1 is therefore unlikely to correspond to the low stoichiometry binding event. For low values of n_{B1}, on the other hand, (the maximum value for) ΔH_{B1} becomes increasingly exothermic. Binding event B1 therefore appears a reasonable candidate for ligand binding to potentially frayed ends: the interaction is more endothermic than the first binding event and therefore has a less favourable entropy of interaction. This fits well with the hypothesis that this binding event corresponds to ligand binding to potentially-frayed strand ends.

Based on an average strand length for poly(dA) of 200 (126-1200 nucleotides, majority 125-350) and the fact that the poly(rU) strands are significantly longer, an estimate for stoichiometry n_{B1} is 0.005 ligands per bp. Although Figure S8 predicts that ΔH_{B1} will not be well defined for this choice, the data was re-analysed using a value of 0.005 for n_{B1} yielding optimised parameter values as illustrated in Figure S9.

Simulated UV-visible titrations

Using the best fit parameters for the model involving two similar binding sites and the model involving one major and one minor binding site, species concentrations were calculated for total concentrations corresponding to those used for the UV-visible titration and these were compared with the experimental data for the UV-visible titration (Figure S10).

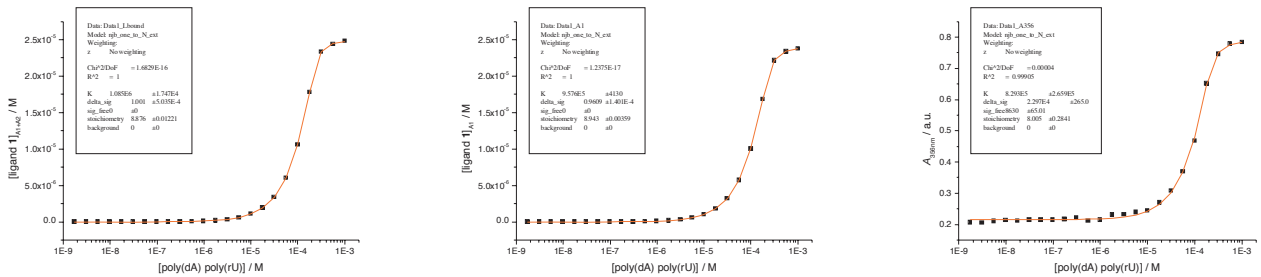


Figure S10: Calculated concentrations [Ligand **1**]$_{A1+A2}$ for the model involving two similar binding sites A1 and A2 (*left*), calculated concentrations [Ligand **1**]$_{A1}$ for the model involving one major and one minor binding site (*middle*) and experimental UV-visible titration (*right*). Lines are fits to a multiple independent binding sites model.

Fitting the simulated data, i.e. assuming a linear UV-visible response with species concentration, to a multiple independent identical binding sites model yields binding constants of 10.9×10^5 M^{-1} and 9.6×10^5 M^{-1} for the simulated data involving two similar binding sites and for the simulated data for a major and a minor binding site, respectively. Both values compare favorably with the experimental binding constant of $(8.3\pm2.7)\times10^5$ M^{-1}. The binding stoichiometry of 8.9 base pairs per ligand as found for both sets of simulated data similarly is in agreement with the value of 8.0 as found for the experimental UV-visible titration. Hence, comparison with the data from the UV-visible titration does not allow us to select one of the models. The observation that simulated data for an equilibrium system involving two different types of binding site can be reproduced extremely well in terms of a model involving a single type of independent binding sites is in line with our previous observations.[2]

Summary

The ITC data do not allow us to distinguish between the two possible combinations of interactions because the two equilibria cannot be satisfactorily separated. This situation resembles the case where data have been collected at a low Wiseman-c. For the case of consecutive binding events, however, the "separability" of binding events depends on the relative Wiseman-c values for the different equilibria and these cannot be optimised through changes in macromolecule concentration

Nevertheless, the relative merits of the two binding models can be compared. In this respect, the model involving two similar binding events suffers from the structural interpretation involving ligand **1** binding to an entire helical turn while affecting binding over a further full helical turn. The model involving a major and a minor binding site, on the other hand, has the advantage that it provides a more realistic structural interpretation of the different binding events in terms of a major binding site involving approximately 10 bp and a minor binding site involving ligand **1** binding to (frayed) ends and gaps.

References

[1] N. J. Buurma, I. Haq, *Methods* **2007**, *42*, 162.
[2] N. J. Buurma, I. Haq, *J. Mol. Biol.* **2008**, *381*, 607.

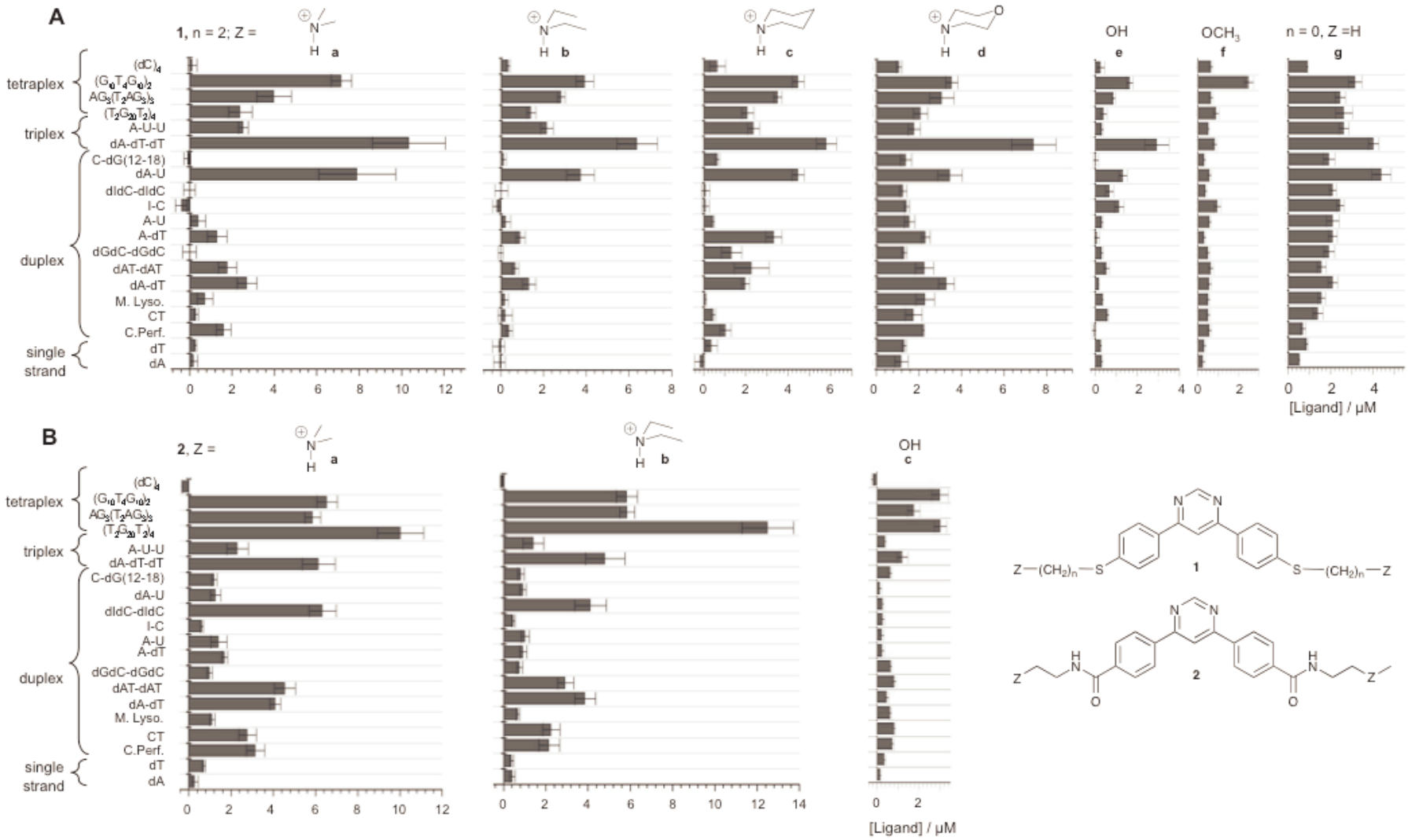


Figure S11. Competition equilibrium dialysis data for compounds (A) thioethers **1** and (B) amides **2**.

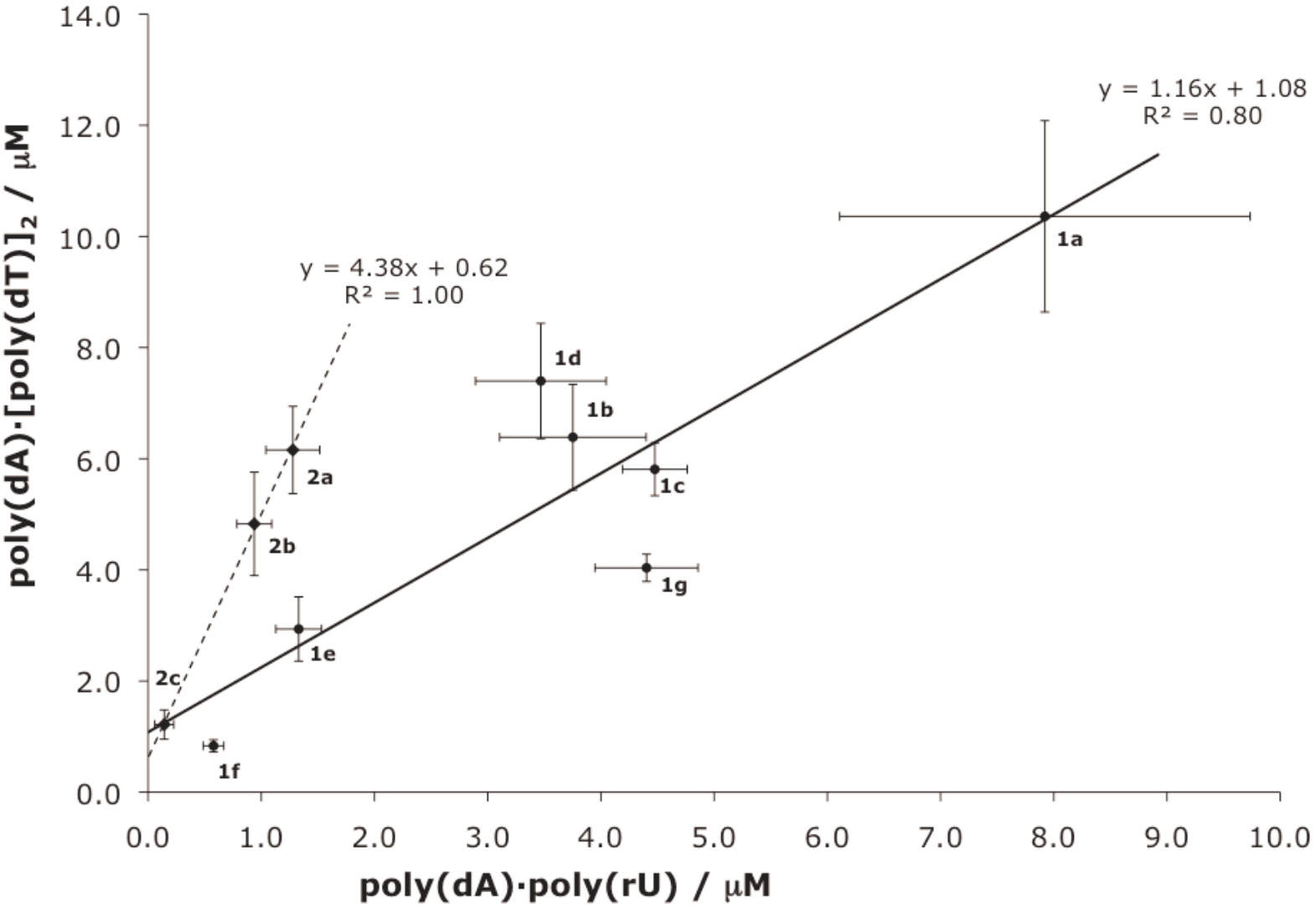


Figure S12. Correlation between DNA triplex and poly(dA)·poly(rU) hybrid duplex binding data from Figure S11.